Who are the Rising Stars in Academia?

Jun Zhang, Zhaolong Ning, Xiaomei Bai, Wei Wang, Shuo Yu, Feng Xia

Dalian University of Technology, Dalian, P.R.China

Introduction

This paper proposes a novel method named ScholarRank to evaluate the scientific impact of rising stars. In our paper, rising stars refer to scholars who are not outstanding among peers or with low research profiles at the beginning stage of their scientific career, but tend to become influential researchers in the future. Our proposed ScholarRank integrates the merits of both statistical indicators and influence calculation algorithms in heterogeneous academic networks. The ScholarRank method considers three factors, which are the citation counts of authors, the mutual influence among coauthors and the mutual reinforce process among different entities in heterogeneous academic networks. The architecture of ScholarRank is shown in Figure 1.

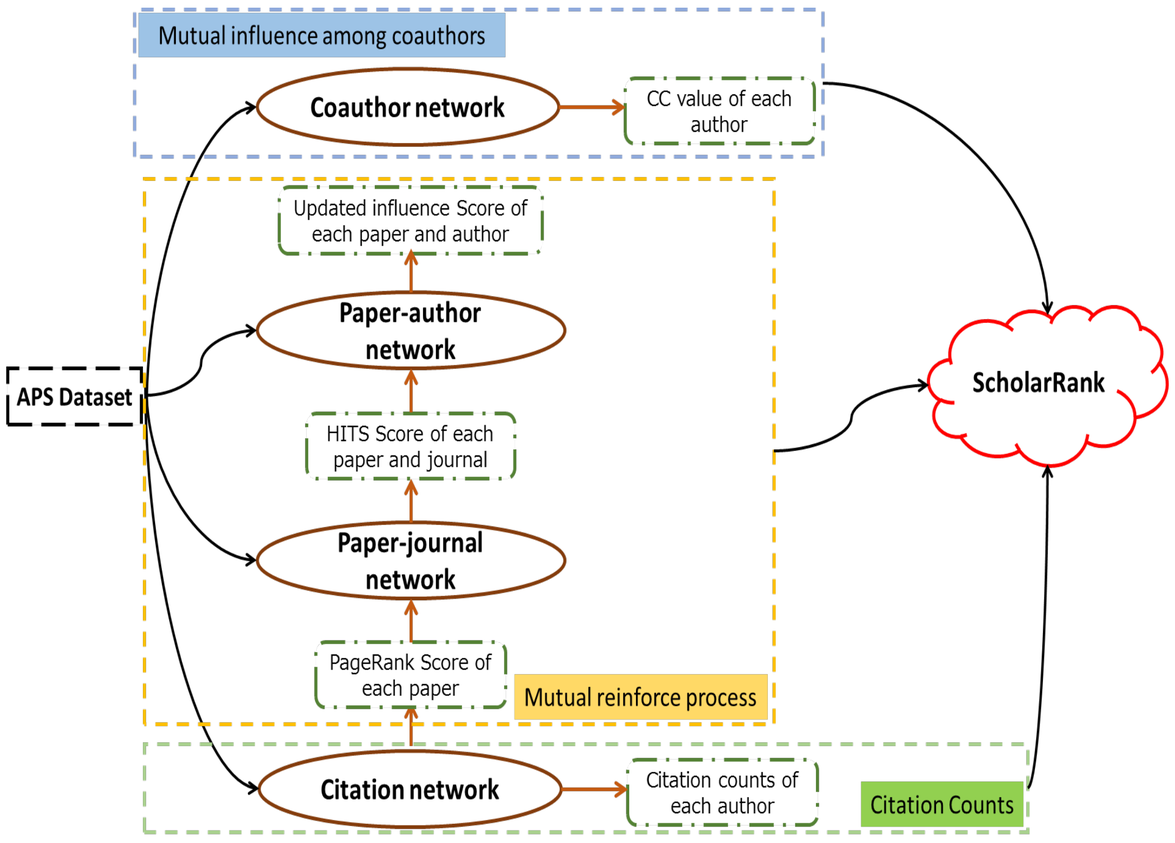

Figure 1. Architecture of ScholarRank

Methods

In order to measure the mutual influence among coauthors, we use the indicator named the caliber of collaboration (CC) to capture the researchers' capacity of collaborating with scholars from diverse backgrounds. The specific method is illustrated as follows.

$$entropy\left(a_i^t\right) = -\sum_{v=1}^{r} w_v^t \log_2\left(w_v^t\right)$$

$$CC\left(a_i\right) = \sum_{t=1}^{u} entropy\left(a_i^t\right)$$

where a_i represents an author, r is the total number of the words in all the institutions' information of a_i's cooperators in year t, and w_v^t is the possibility of word v in all the institutions' information of a_i's cooperators in year t. The value of $CC(a_i)$ is the sum of $entropy\left(a_i^t\right)$ according to specific time intervals, where u refers to the time intervals as we set.

To measure the mutual reinforce process in heterogeneous academic networks, we first construct three sub-networks, i.e. citation network, paper-journal network and paper-author network. The importance of papers is calculated under citation network according to PageRank algorithm. Then we consider the mutual reinforce process between papers and corresponding journals to measure the influence of journals. The HITS algorithm is applied to calculate the influence of journals in paper-journal network. After the calculations in paper-journal network, we then measure the influence of authors also according to HITS algorithm in paper-author network. The following equation is used to calculate the final score of authors:

$$ScholarRank\left(a_i\right) = \frac{1-\alpha-\beta-\delta}{n} + \alpha\frac{CC\left(a_i\right)}{T_{CC}} + \beta\frac{Cita\left(a_i\right)}{T_{Cita}} + \delta\frac{auth\left(a_i\right)}{T_{hyb}}\sum_{j=1}^{\varpi} con\left(a_i^j\right) PR\left(j\right) auth\left(V_j\right)$$

where α, $\beta$and δ are parameters, ϖ is the number of total papers written by author a_i and n is the number of authors in the network. $Cita(a_i)$ is the total citation counts of author a_i and T_{Cita} is the total citation counts of all the authors. T_{CC} is the total CC values of all the authors. $con\left(a_i^j\right)$ means a_i's contribution in paper j and we set it as $1/\theta$ for simplicity, where θ is the order of a_i in paper j. $PR(j)$ is paper j's PagaRank score in citation network, $auth(V_j)$ is the corresponding venue's impact score in paper-journal network and $auth(a_i)$ is the influence score of author a_i in paper-author network. T_{hyb} is the total values of the hybrid results by all the authors.

Simulation and Results

We use datasets from American Physical Society and select authors beginning their scientific careers at the year of 1993. We compare ScholarRank with the following methods to evaluate its effectiveness. In order to validate the performance of our proposed ScholarRank method, we compute the top 10 rising stars' average citation counts in 2013 by our ScholarRank and the above comparison methods. As shown in Figure 2, the ScholarRank achieves the highest average citation counts among all the comparison methods, and it indicates that our proposed ScholarRank can efficiently select top ranking researchers than other methods.

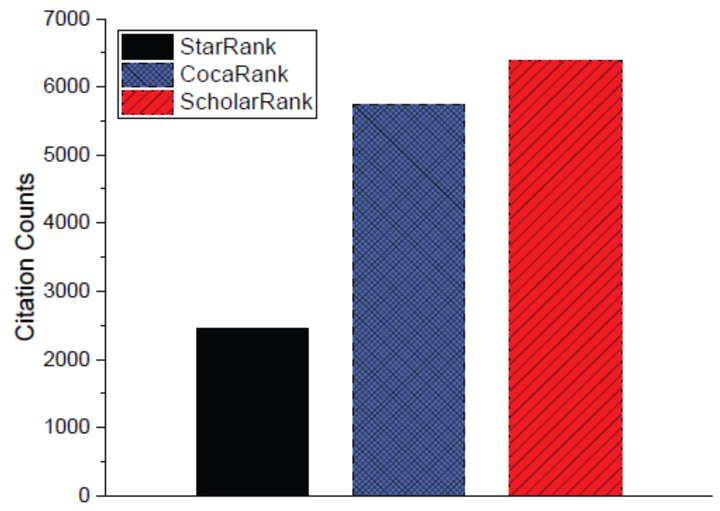

Figure 2. Comparison of average citation counts.

Conclusion

In this paper, we propose the ScholarRank method to evaluate the impact of rising stars, and the experiments on real datasets indicate that our method can find more top ranked rising stars than other methods. In future work, we will test the performance of ScholarRank on more datasets and consider more factors which correlate with the influence of scholars, such as the social relations of scholars, and the download times of papers.